Visualizing Mechanics of Recurrent Neural Networks Through Handwriting Prediction

Pranav A[1]

[1] *20478966, Hong Kong University of Science and Technology*

1 Introduction

Recurrent Neural Networks (RNNs) are commonly used within the sequence based inputs. In the field of Natural Language Processing, the RNNs are used to predict the next word in the sentence. Here, I have presented the paper whether authors have attempt to predict the handwriting strokes based on the previous inputs provided (Carter et al., 2016).

A specialized form of RNNs called Long Short-Term Memory networks (LSTMs) has been used here. Based on the sequence provided, the LSTM tries to predict whether to have a stroke of left or right. In this essay, I will summarize research questions, that authors aim to analyse through visualization of the model.

2 Are LSTMs effective?

This experiment tries to visualize whether LSTMs are appropriate model for the handwriting stroke prediction. LSTMs require step sizes and stroke sizes to output the desired model. These inputs are hyperparameters and thus traditional evaluation through machine learning is futile. Here authors attempt to tune the hyperparameters through visualizing the model output. Of course, these hyperparameters are dependent upon the language and style of writing used. This has been shown in figure 1.

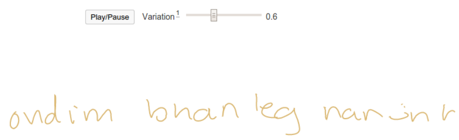

Figure 1: *Tuning hyperparameters like step size and generating the handwriting output.*

3 How to make the model converge?

Edge aggregation is a common technique used in visualization to observe the common paths. Authors aim to use this notion here to analyze the main path of cursive writing by varying the hyperparamters of LSTMs and aggregating the main edge to predict the output of handwriting stroke. This has been shown in figure 2. Through visualization, the user can observe where the model tends to converge (less clutter) and where the

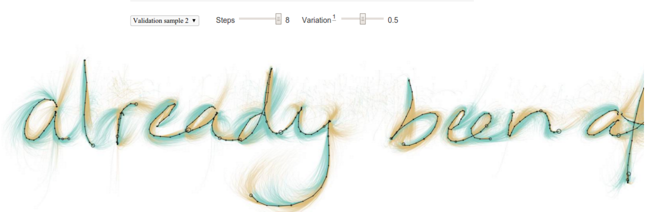

Figure 2: *Here 50 varying outputs have been generated by LSTM by varying step sizes. Edge aggregation technique is applied to visualize the main handwritten paths.*

output is unpredictable (more clutter).

4 How active are the cells of LSTM?

One of the common problems in deep learning is that we mainly think neural networks as a black box algorithm. There has been less research on how different layers work in neural networks. Data visualization of these inputs could be a better idea to see how active these layers or cells.

The model of the authors used had 500 cells. For each stroke of the output, they visualized whether the prediction is leftwards or rightwards. For leftwards intensity, authors choose darker tones of orange progressively. Similarly, for rightwards intensity, authors choose darker tones of green progressively. This experiment has been visualized in figure 3.

Figure 3: *Cell activation visualization of the handwritten input. At a given point, cell can predict right direction (green) or left direction (orange).*

5 How to make the cell activation visualization clearer?

In figure 3, we observed that the activation graph is quite unclear and cluttered. One cannot draw conclusion based on which set of cells are responsible for movement at which type of curve. Is there any way to make this visualization better?

Authors used 1-dimensional t-SNE here to cluster the cells together, which activate approximately together given a certain type of input curve. The has been illustrated in Figure 4.

One can observe that lower set of cells focus on pen-lifts and upper set of cells focus on curvatures.

6 Concluding Remarks

Given the rise of deep learning research, I think these visualization experiments are novel in the way to analyze their inner mechanics. Neural networks are highly effective, but we treat them as a black box and don't

Figure 4: *Cell activation diagram after applying one-dimensional t-SNE. Upper cells indicate curvature changes while lower cells indicate pen lifts.*

analyze how such complex architectures work. This visualize brings the state-of-the-art algorithms in deep learning and visualization together and illustrate how this is applicable on handwriting inputs. I found this visualization in Distill journal which aims at these kind of papers. In my view, machine learning models can be best understood in an user-interface context and this paper succeeds in that. Every paper in AI field should aim to design accessible visualizations like these, so that beginners could understand them better, reproduce them and debug them.

Bibliography

Carter, Shan et al. (2016). “Experiments in Handwriting with a Neural Network”. In: *Distill*. DOI: 10.23915/distill.00004. URL: http://distill.pub/2016/handwriting.